RESEARCH CENTRE

Rennes - Bretagne Atlantique

IN PARTNERSHIP WITH:

CNRS, Université Rennes 1

2021

ACTIVITY REPORT

Project-Team SIROCCO

Analysis representation, compression and communication of visual data

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

DOMAIN

Perception, Cognition and Interaction

THEME

Vision, perception and multimedia interpretation

Contents

Project-Team SIROCCO

Creation of the Project-Team: 2012 January 01

Keywords

Computer sciences and digital sciences

A5. – Interaction, multimedia and robotics

A5.3. – Image processing and analysis

A5.4. – Computer vision

A5.9. – Signal processing

A9. – Artificial intelligence

Other research topics and application domains

B6. – IT and telecom

1 Team members, visitors, external collaborators

Research Scientists

- Christine Guillemot [Team leader, Inria, Senior Researcher, HDR]
- Xiaoran Jiang [Inria, Starting Research Position, until Aug 2021]
- Maja Krivokuca [Inria, Starting Research Position, until Jun 2021]
- Mikael Le Pendu [Inria, Starting Research Position]
- Thomas Maugey [Inria, Researcher]
- Claude Petit [Inria, Researcher, from Sep 2021]
- Aline Roumy [Inria, Senior Researcher, HDR]

Post-Doctoral Fellow

- Anju Jose Tom [Inria]

PhD Students

- Ipek Anil Atalay [Tampere University, from Nov 2021]
- Denis Bacchus [Inria]
- Tom Bachard [Univ de Rennes I, from Oct 2021]
- Nicolas Charpenay [Univ de Rennes I]
- Davi Rabbouni De Carvalho Freitas [Inria, from Sep 2021]
- Simon Evain [Inria, until Feb 2021]
- Rita Fermanian [Inria]
- Patrick Garus [Orange Labs, CIFRE]
- Kai Gu [Inria, from Jun 2021]
- Reda Kaafarani [Inria]
- Brandon Le Bon [Inria]
- Arthur Lecert [Inria]
- Yiqun Liu [Ateme Rennes, CIFRE]
- Xuan Hien Pham [Technicolor, CIFRE]
- Remi Piau [Inria, from Oct 2021]
- Soheib Takhtardeshir [MidSweden University, from Nov 2021]
- Samuel Willingham [Inria, from Jun 2021]

Technical Staff

- Sebastien Bellenous [Inria, Engineer]
- Guillaume Le Guludec [Inria, Engineer]
- Navid Mahmoudian Bidgoli [Inria, Engineer, until Aug 2021]
- Hoai Nam Nguyen [Inria, Engineer, until Aug 2021]
- Jinglei Shi [Inria, Engineer, from Jun 2021]

Interns and Apprentices

- Tom Bachard [Inria, from Feb 2021 until Jul 2021]
- Remi Piau [Inria, from Feb 2021 until Jul 2021]

Administrative Assistant

- Caroline Tanguy [Inria]

2 Overall objectives

2.1 Introduction

Efficient processing, i.e., analysis, storage, access and transmission of visual content, with continuously increasing data rates, in environments which are more and more mobile and distributed, remains a key challenge of the signal and image processing community. New imaging modalities, High Dynamic Range (HDR) imaging, multiview, plenoptic, light fields, 360^o videos, generating very large volumes of data contribute to the sustained need for efficient algorithms for a variety of processing tasks.

Building upon a strong background on signal/image/video processing and information theory, the goal of the SIROCCO team is to design mathematically founded tools and algorithms for visual data analysis, modeling, representation, coding, and processing, with for the latter area an emphasis on inverse problems related to super-resolution, view synthesis, HDR recovery from multiple exposures, denoising and inpainting. Even if 2D imaging is still within our scope, the goal is to give a particular attention to HDR imaging, light fields, and 360^o videos. The project-team activities are structured and organized around the following inter-dependent research axes:

- Visual data analysis
- Signal processing and learning methods for visual data representation and compression
- Algorithms for inverse problems in visual data processing
- User-centric compression.

While aiming at generic approaches, some of the solutions developed are applied to practical problems in partnership with industry (InterDigital, Ateme, Orange) or in the framework of national projects. The application domains addressed by the project are networked visual applications taking into account their various requirements and needs in terms of compression, of network adaptation, of advanced functionalities such as navigation, interactive streaming and high quality rendering.

2.2 Visual Data Analysis

Most visual data processing problems require a prior step of data analysis, of discovery and modeling of correlation structures. This is a pre-requisite for the design of dimensionality reduction methods, of compact representations and of fast processing techniques. These correlation structures often depend on the scene and on the acquisition system. Scene analysis and modeling from the data at hand is hence also part of our activities. To give examples, scene depth and scene flow estimation is a cornerstone of many approaches in multi-view and light field processing. The information on scene geometry helps constructing representations of reduced dimension for efficient (e.g. in interactive time) processing of new imaging modalities (e.g. light fields or 360^o videos).

2.3 Signal processing and learning methods for visual data representation and compression

Dimensionality reduction has been at the core of signal and image processing methods, for a number of years now, hence have obviously always been central to the research of Sirocco. These methods encompass sparse and low-rank models, random low-dimensional projections in a compressive sensing framework, and graphs as a way of representing data dependencies and defining the support for learning and applying signal de-correlating transforms. The study of these models and signal processing tools is even more compelling for designing efficient algorithms for processing the large volumes of high-dimensionality data produced by novel imaging modalities. The models need to be adapted to the data at hand through learning of dictionaries or of neural networks. In order to define and learn local low-dimensional or sparse models, it is necessay to capture and understand the underlying data geometry, e.g. with the help of manifolds and manifold clustering tools. It also requires exploiting the scene geometry with the help of disparity or depth maps, or its variations in time via coarse or dense scene flows.

2.4 Algorithms for inverse problems in visual data processing

Based on the above models, besides compression, our goal is also to develop algorithms for solving a number of inverse problems in computer vision. Our emphasis is on methods to cope with limitations of sensors (e.g. enhancing spatial, angular or temporal resolution of captured data, or noise removal), to synthesize virtual views or to reconstruct (e.g. in a compressive sensing framework) light fields from a sparse set of input views, to recover HDR visual content from multiple exposures, and to enable content editing (we focus on color transfer, re-colorization, object removal and inpainting). Note that view synthesis is a key component of multiview and light field compression. View synthesis is also needed to support user navigation and interactive streaming. It is also needed to avoid angular aliasing in some post-capture processing tasks, such as re-focusing, from a sparse light field. Learning models for the data at hand is key for solving the above problems.

2.5 User-centric compression

The ever-growing volume of image/video traffic motivates the search for new coding solutions suitable for band and energy limited networks but also space and energy limited storage devices. In particular, we investigate compression strategies that are adapted to the users needs and data access requests in order to meet all these transmission and/or storage constraints. Our first goal is to address theoretical issues such as the information theoretical bounds of these compression problems. This includes compression of a database with random access, compression with interactivity, and also data repurposing that takes into account the users needs and user data perception. A second goal is to construct practical coding for all these problems.

3 Research program

3.1 Introduction

The research activities on analysis, compression and communication of visual data mostly rely on tools and formalisms from the areas of statistical image modeling, of signal processing, of machine learning, of coding and information theory. Some of the proposed research axes are also based on scientific foundations of computer vision (e.g. multi-view modeling and coding). We have limited this section to some tools which are central to the proposed research axes, but the design of complete compression and communication solutions obviously rely on a large number of other results in the areas of motion analysis, transform design, entropy code design, etc which cannot be all described here.

3.2 Data Dimensionality Reduction

Keywords: Manifolds, graph-based transforms, compressive sensing.

Dimensionality reduction encompasses a variety of methods for low-dimensional data embedding, such as sparse and low-rank models, random low-dimensional projections in a compressive sensing framework, and sparsifying transforms including graph-based transforms. These methods are the cornerstones of many visual data processing tasks (compression, inverse problems).

Sparse representations, *compressive sensing*, and *dictionary learning* have been shown to be powerful tools for efficient processing of visual data. The objective of *sparse representations* is to find a sparse approximation of a given input data. In theory, given a dictionary matrix $A \in \mathbb{R}^{m \times n}$, and a data $\mathbf{b} \in \mathbb{R}^m$ with $m << n$ and A is of full row rank, one seeks the solution of $\min\{\|\mathbf{x}\|_0 : A\mathbf{x} = \mathbf{b}\}$, where $\|\mathbf{x}\|_0$ denotes the ℓ_0 norm of $\mathbf{x}$, i.e. the number of non-zero components in $\mathbf{x}$. A is known as the dictionary, its columns a_j are the atoms, they are assumed to be normalized in Euclidean norm. There exist many solutions x to $Ax = b$. The problem is to find the sparsest solution x, i.e. the one having the fewest nonzero components. In practice, one actually seeks an approximate and thus even sparser solution which satisfies $\min\{\|\mathbf{x}\|_0 : \|A\mathbf{x} - \mathbf{b}\|_p \leq \rho\}$, for some $\rho \geq 0$, characterizing an admissible reconstruction error.

The recent theory of *compressed sensing*, in the context of discrete signals, can be seen as an effective dimensionality reduction technique. The idea behind compressive sensing is that a signal can be accurately recovered from a small number of linear measurements, at a rate much smaller than what is commonly prescribed by the Shannon-Nyquist theorem, provided that it is sparse or compressible in a known basis. Compressed sensing has emerged as a powerful framework for signal acquisition and sensor design, with a number of open issues such as learning the basis in which the signal is sparse, with the help of dictionary learning methods, or the design and optimization of the sensing matrix. The problem is in particular investigated in the context of light fields acquisition, aiming at novel camera design with the goal of offering a good trade-off between spatial and angular resolution.

While most image and video processing methods have been developed for cartesian sampling grids, new imaging modalities (e.g. point clouds, light fields) call for representations on irregular supports that can be well represented by *graphs*. Reducing the dimensionality of such signals require designing novel transforms yielding compact signal representation. One example of transform is the Graph Fourier transform whose basis functions are given by the eigenvectors of the graph Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is a diagonal degree matrix whose i^{th} diagonal element is equal to the sum of the weights of all edges incident to the node i, and $\mathbf{A}$ the adjacency matrix. The eigenvectors of the Laplacian of the graph, also called Laplacian eigenbases, are analogous to the Fourier bases in the Euclidean domain and allow representing the signal residing on the graph as a linear combination of eigenfunctions akin to Fourier Analysis. This transform is particularly efficient for compacting smooth signals on the graph. The problems which therefore need to be addressed are (i) to define graph structures on which the corresponding signals are smooth for different imaging modalities and (ii) the design of transforms compacting well the signal energy with a tractable computational complexity.

3.3 Deep neural networks

Keywords: Autoencoders, Neural Networks, Recurrent Neural Networks.

From dictionary learning which we have investigated a lot in the past, our activity is now evolving towards deep learning techniques which we are considering for dimensionality reduction. We address the problem of unsupervised learning of transforms and prediction operators that would be optimal in terms of energy compaction, considering autoencoders and neural network architectures.

An autoencoder is a neural network with an encoder g_e, parametrized by θ, that computes a representation Y from the data X, and a decoder g_d, parametrized by ϕ, that gives a reconstruction $\hat{X}$ of X (see Figure below). Autoencoders can be used for dimensionality reduction, compression, denoising. When it is used for compression, the representation need to be quantized, leading to a quantized representation $\hat{Y} = Q(Y)$ (see Figure below). If an autoencoder has fully-connected layers, the architecture, and the number of parameters to be learned, depends on the image size. Hence one autoencoder has to be trained per image size, which poses problems in terms of genericity. To avoid this limitation, architectures

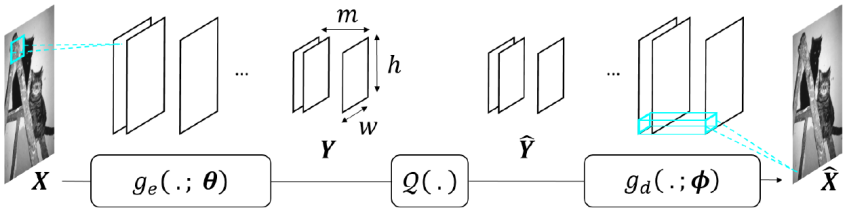


Figure 1: Illustration of an autoencoder.

without fully-connected layer and comprising instead convolutional layers and non-linear operators, forming convolutional neural networks (CNN) may be preferrable. The obtained representation is thus a set of so-called feature maps.

The other problems that we address with the help of neural networks are scene geometry and scene flow estimation, view synthesis, prediction and interpolation with various imaging modalities. The problems are posed either as supervised or unsupervised learning tasks. Our scope of investigation includes autoencoders, convolutional networks, variational autoencoders and generative adversarial networks (GAN) but also recurrent networks and in particular Long Short Term Memory (LSTM) networks. Recurrent neural networks attempting to model time or sequence dependent behaviour, by feeding back the output of a neural network layer at time t to the input of the same network layer at time t+1, have been shown to be interesting tools for temporal frame prediction. LSTMs are particular cases of recurrent networks made of cells composed of three types of neural layers called gates.

Deep neural networks have also been shown to be very promising for solving inverse problems (e.g. super-resolution, sparse recovery in a compressive sensing framework, inpainting) in image processing. Variational autoencoders, generative adversarial networks (GAN), learn, from a set of examples, the latent space or the manifold in which the images, that we search to recover, reside. The inverse problems can be re-formulated using a regularization in the latent space learned by the network. For the needs of the regularization, the learned latent space may need to verify certain properties such as preserving distances or neighborhood of the input space, or in terms of statistical modeling. GANs, trained to produce images that are plausible, are also useful tools for learning texture models, expressed via the filters of the network, that can be used for solving problems like inpainting or view synthesis.

3.4 Coding theory

Keywords: OPTA limit (Optimum Performance Theoretically Attainable), Rate allocation, Rate-Distortion optimization, lossy coding, joint source-channel coding multiple description coding, channel modelization, oversampled frame expansions, error correcting codes..

Source coding and channel coding theory [1] is central to our compression and communication activities, in particular to the design of entropy codes and of error correcting codes. Another field in coding theory which has emerged in the context of sensor networks is Distributed Source Coding (DSC). It refers to the compression of correlated signals captured by different sensors which do not communicate between themselves. All the signals captured are compressed independently and transmitted to a central base station which has the capability to decode them jointly. DSC finds its foundation in the seminal Slepian-Wolf [2] (SW) and Wyner-Ziv [3] (WZ) theorems. Let us consider two binary correlated sources X and Y. If the two coders communicate, it is well known from Shannon's theory that the minimum lossless rate for X and Y is given by the joint entropy $H(X,Y)$. Slepian and Wolf have established in 1973 that this lossless compression rate bound can be approached with a vanishing error probability for long sequences, even if the two sources are coded separately, provided that they are decoded jointly and that their correlation is known to both the encoder and the decoder.

In 1976, Wyner and Ziv considered the problem of coding of two correlated sources X and Y, with respect to a fidelity criterion. They have established the rate-distortion function $R*_{X|Y}(D)$ for the case where the side information Y is perfectly known to the decoder only. For a given target distortion D, $R*_{X|Y}(D)$ in general verifies $R_{X|Y}(D) \leq R*_{X|Y}(D) \leq R_X(D)$, where $R_{X|Y}(D)$ is the rate required to encode X if Y is available to both the encoder and the decoder, and R_X is the minimal rate for encoding X without SI. These results give achievable rate bounds, however the design of codes and practical solutions for compression and communication applications remain a widely open issue.

4 Application domains

4.1 Overview

The application domains addressed by the project are:

- Compression with advanced functionalities of various imaging modalities
- Networked multimedia applications taking into account needs in terms of user and network adaptation (e.g., interactive streaming, resilience to channel noise)
- Content editing, post-production, and computational photography.

4.2 Compression of emerging imaging modalities

Compression of visual content remains a widely-sought capability for a large number of applications. This is particularly true for mobile applications, as the need for wireless transmission capacity will significantly increase during the years to come. Hence, efficient compression tools are required to satisfy the trend towards mobile access to larger image resolutions and higher quality. A new impulse to research in video compression is also brought by the emergence of new imaging modalities, e.g. high dynamic range (HDR) images and videos (higher bit depth, extended colorimetric space), light fields and omni-directional imaging.

Different video data formats and technologies are envisaged for interactive and immersive 3D video applications using omni-directional videos, stereoscopic or multi-view videos. The "omni-directional video" set-up refers to 360-degree view from one single viewpoint or spherical video. Stereoscopic video is composed of two-view videos, the right and left images of the scene which, when combined, can recreate the depth aspect of the scene. A multi-view video refers to multiple video sequences captured by multiple video cameras and possibly by depth cameras. Associated with a view synthesis method, a multi-view video allows the generation of virtual views of the scene from any viewpoint. This property can be used in a large diversity of applications, including Three-Dimensional TV (3DTV), and Free Viewpoint Video (FVV). In parallel, the advent of a variety of heterogeneous delivery infrastructures has given

[1] T. M. Cover and J. A. Thomas, Elements of Information Theory, Second Edition, July 2006.

[2] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources." IEEE Transactions on Information Theory, 19(4), pp. 471-480, July 1973.

[3] A. Wyner and J. Ziv, "The rate-distortion function for source coding ith side information at the decoder." IEEE Transactions on Information Theory, pp. 1-10, January 1976.

momentum to extensive work on optimizing the end-to-end delivery QoS (Quality of Service). This encompasses compression capability but also capability for adapting the compressed streams to varying network conditions. The scalability of the video content compressed representation and its robustness to transmission impairments are thus important features for seamless adaptation to varying network conditions and to terminal capabilities.

4.3 Networked visual applications

Free-viewpoint Television (FTV) is a system for watching videos in which the user can choose its viewpoint freely and change it at anytime. To allow this navigation, many views are proposed and the user can navigate from one to the other. The goal of FTV is to propose an immersive sensation without the disadvantage of Three-dimensional television (3DTV). With FTV, a look-around effect is produced without any visual fatigue since the displayed images remain 2D. However, technical characteristics of FTV are large databases, huge numbers of users, and requests of subsets of the data, while the subset can be randomly chosen by the viewer. This requires the design of coding algorithms allowing such a random access to the pre-encoded and stored data which would preserve the compression performance of predictive coding. This research also finds applications in the context of Internet of Things in which the problem arises of optimally selecting both the number and the position of reference sensors and of compressing the captured data to be shared among a high number of users.

Broadband fixed and mobile access networks with different radio access technologies have enabled not only IPTV and Internet TV but also the emergence of mobile TV and mobile devices with internet capability. A major challenge for next internet TV or internet video remains to be able to deliver the increasing variety of media (including more and more bandwidth demanding media) with a sufficient end-to-end QoS (Quality of Service) and QoE (Quality of Experience).

4.4 Editing, post-production and computational photography

Editing and post-production are critical aspects in the audio-visual production process. Increased ways of "consuming" visual content also highlight the need for content repurposing as well as for higher interaction and editing capabilities. Content repurposing encompasses format conversion (retargeting), content summarization, and content editing. This processing requires powerful methods for extracting condensed video representations as well as powerful inpainting techniques. By providing advanced models, advanced video processing and image analysis tools, more visual effects, with more realism become possible. Our activies around light field imaging also find applications in computational photography which refers to the capability of creating photographic functionalities beyond what is possible with traditional cameras and processing tools.

5 Social and environmental responsibility

No social or environmental responsibility.

6 Highlights of the year

This year has seen the start of

- The H2020 Marie Sklodowska-Curie Innovative Training Network Plenoptima dealing with plenoptic Imaging, in collaboration with Tampere University, MidSweden University, and the Technical University of Berlin.
- The exploratory action DARE (Data Repurposing)
- The CominLabs MOVE (Mature Omnidirectional Video Exploration) project.
- Inria Start-up Studio project called Anax.

7 New software and platforms

This section describes the new software developed in the year 2021 as well as the datasets created and the platform under development.

7.1 New software

7.1.1 Interactive compression for omnidirectional images and texture maps of 3D models

Keywords: Image compression, Random access

Functional Description: This code implements a new image compression algorithm that allows to navigate within a static scene. To do so, the code provides access in the compressed domain to any block and therefore allows extraction of any subpart of the image. This codec implements this interactive compression for two image modalities: omnidirectional images and texture maps of 3D models. For omnidirectional images the input is a 2D equirectangular projection of the 360 image. The output is the image seen in the viewport. For 3D models, the input is a texture map and the 3D mesh. The output is also the image seen in the viewport.

The code consists of three parts: (A) an offline encoder (B) an online bit extractor and (C) a decoder. The offline encoder (i) partitions the image into blocks, (ii) optimizes the positions of the access blocks, (iii) computes a set of geometry aware predictions for each block (to cover all possible navigation paths), (iv) implements transform quantization for all blocks and their predictions, and finally (v) evaluates the encoding rates. The online bit extractor (Part B) first computes the optimal and geometry aware scanning order. Then it extracts in the bitstream, the sufficient amount of information to allow the decoding of the requested blocks. The last part of the code is the decoder (Part C). The decoder reconstructs the same scanning order as the one computed at the online bit extractor. Then, the blocks are decoded (inverse transform, geometry aware predictions, ...) and reconstructed. Finally the image in the viewport is generated.

Authors: Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy

Contact: Aline Roumy

7.1.2 Rate-distortion optimized motion estimation for on the sphere compression of 360 videos

Keywords: Image compression, Omnidirectional video

Functional Description: This code implements a new video compression algorithm for omnidirectional (360°) videos. The main originality of this algorithm is that the compression is performed directly on the sphere. First, it saves computational complexity as it avoids to project the sphere onto a 2D map, as classically done. Second, and more importantly, it allows to achieve a better rate-distortion tradeoff, since neither the visual data nor its domain are distorted. This code implements an extension from still images to videos of the on-the-sphere compression for omnidirectional still images. In particular, it implements a novel rate-distortion optimized motion estimation algorithm to perform motion compensation. The optimization is performed among a set of existing motion models and a novel motion model called tangent-linear+t. Moreover, a finer search pattern, called spherical-uniform, is also implemented for the motion parameters, which leads to a more accurate block prediction. The novel algorithm leads to rate-distortion gains compared to methods based on a unique motion model.

More precisely, this algorithm performs inter-prediction and contains (i) a motion estimation, (ii) a motion compensation and computation of a residue, (iii) the encoding of this residue, (iv) estimation of rate and distortion. Several visualization tools are also provided such as rate-distortion curves, but also search pattern centers, and motion field.

Authors: Alban Marie, Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy

Contact: Aline Roumy

7.1.3 GRU_DDLF

Name: Gated Recurrent Unit and Deep Decoder based generative model for Light Field compression

Keywords: Light fields, Neural networks, Deep learning, Compression

Functional Description: This code implements a neural network-based light field representation and compression scheme The code is developped in python, based on the pytorch deep learning framework..The proposed network is based on both a generative model that aims at modeling the spatial information that is static, i.e., found in all light field views, and on a convolutional Gated Recurrent Unit (ConvGRU) that is used to model variations between blocks of angular views. The network is untrained in the sense that it is learned only on the light field to be processed, without any additional training data. The network weights can be considered as a representation of the input light field. The compactness of this representation depends on the number of weights or network parameters, but not only. It also depends on the number of bits needed to accurately quantize each weight. Our network is thus learned using a strategy that takes into account weight quantization, in order to minimize the effect of weight quantization noise on the light field reconstruction quality. The weights of the different filters are optimized end-to-end in order to optimize the image reconstruction quality for a given number of quantization bits per weight.

Contact: Christine Guillemot

Participants: Xiaoran Jiang, Jinglei Shi, Christine Guillemot

7.1.4 FPFR+

Name: Fused Pixel and Feature based Reconstruction for view synthesis and temporal interpolation

Keywords: Light fields, View synthesis, Temporal interpolation, Neural networks

Functional Description: This code implements a deep residual architecture that can be used both for synthesizing high quality angular views in light fields and temporal frames in classical videos. The code is developped in python, based on the tensorflow deep learning framework. The proposed framework consists of an optical flow estimator optimized for view synthesis, a trainable feature extractor and a residual convolutional network for pixel and feature-based view reconstruction. Among these modules, the fine-tuning of the optical flow estimator specifically for the view synthesis task yields scene depth or motion information that is well optimized for the targeted problem. In cooperation with the end-to-end trainable encoder, the synthesis block employs both pixel-based and feature-based synthesis with residual connection blocks, and the two synthesized views are fused with the help of a learned soft mask to obtain the final reconstructed view.

Contact: Christine Guillemot

Participants: Jinglei Shi, Xiaoran Jiang, Christine Guillemot

7.1.5 PnP-A

Name: Plug-and-play algorithms

Keywords: Algorithm, Inverse problem, Deep learning, Optimization

Functional Description: The software is a framework for solving inverse problems using so-called "plug-and-play" algorithms in which the regularisation is performed with an external method such as a denoising neural network. The framework also includes the possibility to apply preconditioning to the algorithms for better performances. The code is developped in python, based on the pytorch deep learning framework, and is designed in a modular way in order to combine each inverse problem (e.g. image completion, interpolation, demosaicing, denoising, ...) with different algorithms (e.g. ADMM, HQS, gradient descent), diifferent preconditioning methods, and different denoising neural networks.

Contact: Christine Guillemot

Participants: Mikael Le Pendu, Christine Guillemot

7.1.6 SIUPPA

Name: Stochastic Implicit Unrolled Proximal Point Algorithm

Keywords: Inverse problem, Optimization, Deep learning

Functional Description: This code implements a stochastic implicit unrolled proximal point method, where an optimization problem is defined for each iteration of the unrolled ADMM scheme, with a learned regularizer. The code is developped in python, based on the pytorch deep learning framework. The unrolled proximal gradient method couples an implicit model and a stochastic learning strategy. For each backpropagation step, the weights are updated from the last iteration as in the Jacobian-Free Backpropagation Implicit Networks, but also from a randomly selected set of unrolled iterations. The code includes several applications, namely denoising, super-resolution, deblurring and demosaicing.

Contact: Christine Guillemot

Participants: Brandon Le Bon, Mikael Le Pendu, Christine Guillemot

7.1.7 DeepLFCam

Name: Deep Light Field Acquisition Using Learned Coded Mask Distributions for Color Filter Array Sensors

Keywords: Light fields, Deep learning, Compressive sensing

Functional Description: This code implements a deep light field acquisition method using learned coded mask distributions for color filter array sensors. The code describes and allows the execution of an algorithm for dense light field reconstruction using a small number of simulated monochromatic projections of the light field. Those simulated projections consist of the composition of: - the filtering of a light field by a color coded mask placed between the aperture plane and the sensor plane, performing both angular and spectral multiplexing, - a color filtering array performing color multiplexing, - a monochromatic sensor. This composition of filtering, sampling, projection, is modeled as linear projection operator. Those measurements, along with the 'modulation field', or 'shield field' corresponding to the acquisition device (i.e. the light field corresponding to the the transformation of a uniformly white light field, known to completelly characterize the projection operator), are then feed to a deep convolutional residual network to reconstruct the original light field. The provided implementation is meant to be used with Tensorflow 2.1.

Publication: leguludec:hal-03203347

Contact: Guillaume Le Guludec

Participants: Guillaume Le Guludec, Christine Guillemot

7.1.8 DeepULFCam

Name: Deep Unrolling for Light Field Compressed Acquisition using Coded Masks

Keywords: Light fields, Optimization, Deep learning, Compressive sensing

Functional Description: This code describes and allows the execution of an algorithm for dense light field reconstruction using a small number of simulated monochromatic projections of the light field. Those simulated projections consist of the composition of: - the filtering of a light field by a color coded mask placed between the aperture plane and the sensor plane, performing both angular and spectral multiplexing, - a color filtering array performing color multiplexing, - a monochromatic

sensor. The composition of these filterings/projects is modeled as linear projection operator. The light field is then reconstructed by performing the 'unrolling' of an interative reconstruction algorithm (namely the HQS 'half-quadratic splitting' algorithm, a variant of ADMM, an optimization algorithm, applied to the solving a regularized least-squares problem) into a deep convolutional neural network. The algorithm makes use of the structure of the projection operator to efficiently solve the quadratic data-fidelty minimization sub-problem in closed form. This code is designed to be compatible with Python 3.7 and Tensorflow 2.3. Other required dependencies are: Pillow, PyYAML.

Contact: Guillaume Le Guludec

Participants: Guillaume Le Guludec, Christine Guillemot

7.1.9 OSLO: On-the-Sphere Learning for Omnidirectional images

Keywords: Omnidirectional image, Machine learning, Neural networks

Functional Description: This code implements a deep convolutional neural network for omnidirectional images. The approach operates directly on the sphere, without the need to project the data on a 2D image. More specifically, from the sphere pixelization, called healpix, the code implements a convolution operation on the sphere that keeps the orientation (north/south/east/west). This convolution has the same complexity as a classical 2D planar convolution. Moreover, the code implements stride, iterative aggregation, and pixel shuffling in the spherical domain. Finally, image compression is implemented as an application of this on-the-sphere CNN.

Authors: Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy

Contact: Aline Roumy

7.2 New platforms

7.2.1 Acquisition of multi-view sequences for Free viewpoint Television

Participants: Laurent Guillo, Thomas Maugey.

The scientific and industrial community is nowadays exploring new multimedia applications using 3D data (beyond stereoscopy). In particular, Free Viewpoint Television (FTV) has attracted much attention in the recent years. In those systems, user can choose in real time its view angle from which he wants to observe the scene. Despite the great interest for FTV, the lack of realistic and ambitious datasets penalizes the research effort. The acquisition of such sequences is very costly in terms of hardware and working effort, which explains why no multi-view videos suitable for FTV has been proposed yet.

In the context of the project ADT ATeP 2016-2018 (funded by Inria), such datasets were acquired and some calibration tools have been developed. First 40 omnidirectional cameras and their associated equipments have been acquired by the team (thanks to Rennes Metropole funding). We have first focused on the calibration of this camera, *i.e.*, the development of the relationship between a 3D point and its projection in the omnidirectional image. In particular, we have shown that the unified spherical model fits the acquired omnidirectional cameras. Second, we have developed tools to calibrate the cameras in relation to each other. Finally, we have made a capture of 3 multiview sequences that have been made available to the community via a public web site.

7.2.2 CLIM processing toolbox

Participants: Christine Guillemot, Laurent Guillo.

As part of the ERC Clim project, the EPI Sirocco is developing a light field processing toolbox. The toolbox and libraries are developed in C++ and the graphical user interface relies on Qt. As input data, this tool accepts both sparse light fields acquired with High Density Camera Arrays (HDCA) and denser light fields captured with plenoptic cameras using microlens arrays (MLA). At the time of writing, in addition to some simple functionalities, such as re-focusing, change of viewpoints, with different forms of visualization, the toolbox integrates more advanced tools for scene depth estimation from sparse and dense light fields, for super-ray segmentation and scene flow estimation, and for light field denoising and angular interpolation using anisotropic diffusion in the 4D ray space. The toolbox is now being interfaced with the C/C++ API of the tensorflow platform, in order to execute deep models developed in the team for scene depth and scene flow estimation, view synthesis, and axial super-resolution.

7.2.3 ADT: Interactive Coder of Omnidirectional Videos

Participants: Sebastien Bellenous, Navid Mahmoudian Bidgoli, Thomas Maugey.

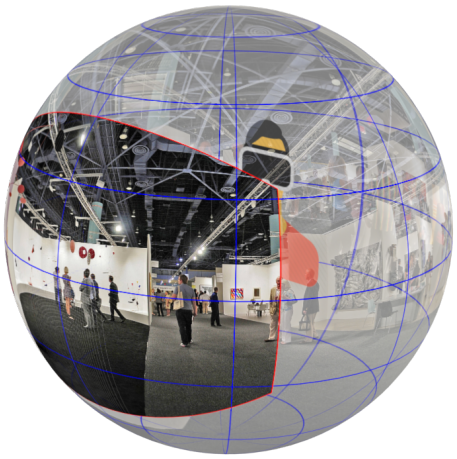


Figure 2: By nature, a video 360° cannot be seen entirely, because it covers all the direction. The developed coding scheme will aim at 1) transmitting only what is necessary and 2) bringing no transmission rate overhead.

In the Intercom project, we have studied the impact of interactivity on the coding performance. We have, for example, tackled the following problem: is it possible to compress a 360° video (as shown in Fig.2) once for all, and then partly extract and decode what is needed for a user navigation, while, keeping good compression performance? After having derived the achievable theoretical bounds, we have built a new omnidirectional video coder. This original architecture enables to reduce significantly the cost of interactivity compared to the conventional video coders. The project ICOV proposes to start from this promising proof of concept and to develop a complete and well specified coder that is aimed to be shared with the community.

8 New results

8.1 Visual Data Analysis

Keywords: Scene depth, Scene flows, 3D modeling, Light-fields, camera design, 3D point clouds.

8.1.1 Deep Light Field Acquisition Using Learned Coded Mask Distributions for Color Filter Array Sensors

Participants: Christine Guillemot, Guillaume Le Guludec.

Compressive light field photography enables light field acquisition using a single sensor by utilizing a color coded mask. This approach is very cost effective since consumer-level digital cameras can be turned into a light field camera by simply placing a coded mask between the sensor and the aperture plane and solving an inverse problem to obtain an estimate of the original light field. While in the past years, we developed solutions based on signal processing methods [8], in 2021 we have developed a deep learning architecture for compressive light field acquisition using a color coded mask and a sensor with Color Filter Array (CFA) [10], in line with the multi-mask camera model we proposed in [12]. Unlike previous methods where a fixed mask pattern is used, our deep network learns the optimal distribution of the color coded mask pixels. The proposed solution enables end-to-end learning of the color-coded mask distribution and the reconstruction network, taking into account the sensor CFA. Consequently, the resulting network can efficiently perform joint demosaicing and light field reconstruction of images acquired with color-coded mask and a CFA sensor. Compared to previous methods based on deep learning with monochrome sensors, as well as traditional compressive sensing approaches using CFA sensors, we obtain superior color reconstruction of the light fields.

We have also presented an efficient and mathematically grounded deep learning model to reconstruct a light field from a set of measurements obtained using a color-coded mask and a color filter array (CFA). Following the promising trend of unrolling optimization algorithms with learned priors, we formulate our task of light field reconstruction as an inverse problem and derive a principled deep network architecture from this formulation. We also introduce a closed-form extraction of information from the acquisition, while similar methods found in the recent literature systematically use an approximation. Compared to similar deep learning methods, we show that our approach allows for a better reconstruction quality. We have further shown that our approach is robust to noise using realistic simulations of the sensing acquisition process.

8.1.2 Compressive HDR Light Field Camera with Multiple ISO Sensors

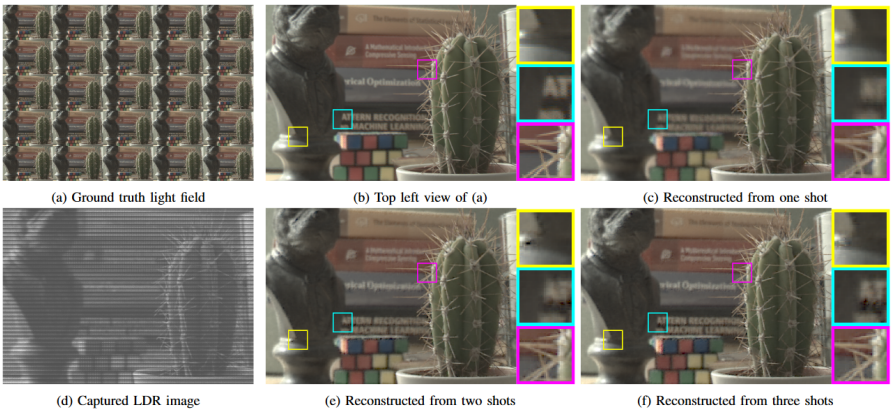
Participants: Christine Guillemot, Hoai Nam Nguyen.

The problem of HDR light field acquisition using a 2D sensor remains an open and challenging problem despite the recent advances in both 2D HDR imaging and compressive LDR light field acquisition. The main challenge here is indeed the reconstruction of an HDR light field from a single LDR image recorded on a monochrome sensor equipped with a Color Filter Array (CFA). This single monochrome image hence should encode not only the HDR information of the scene, but also angular and spectral measurements of the light field.

To address this problem, in collaboration with the Univ. of Linkoping, we have introduced a novel framework for compressive capture of HDR light fields combining multiple ISO photography with mask-based coded projection techniques [11]. The approach builds upon the multi-mask camera model we proposed in [12], and based on a main lens, a multi-ISO sensor and a coded mask located in the optical path between the main lens and the sensor. The mask projects coded spatio-angular information of the light field onto the 2D sensor. Hence, our compressive HDR light field imaging framework captures a coded image with a varying per pixel gain encoding the scene. The sensor image captured through the mask, the varying per pixel gain, and the CFA, encodes spatial, angular, and color intensity variations in the scene. This coded projection image compresses the incident scene radiance information such that the full HDR light field can be recovered as a tractable inverse problem. The model encompasses different acquisition scenarios with different ISO patterns and gains. Moreover, we assume that the sensor has a built-in color filter array (CFA), making our design more suitable for consumer-level cameras.

We, in parallel, developed a novel joint spatio-angular-HDR reconstruction algorithm using a trained dictionary specifically designed for HDR light field reconstruction. The joint reconstruction includes a confidence matrix based on the pixel intensity and acquisition noise, effectively performing denoising

as an integral part in the reconstruction. The reconstruction algorithm actually jointly performs color demosaicing, light field angular information recovery, HDR reconstruction, and denoising from the multi-ISO measurements formed on the sensor (see an illustration of some results in Fig.(3).



(a) Ground truth light field (b) Top left view of (a) (c) Reconstructed from one shot

(d) Captured LDR image (e) Reconstructed from two shots (f) Reconstructed from three shots

Figure 3: LDR sensor image and the reconstruction results of the Cactus light field assuming a CFA sensor with four ISO values (100, 400, 800 and 1600) using one, two, and three shots.

We have also created two HDR light field data sets: one synthetic data set created using the Blender rendering software with two baselines, and a real light field data set created from the fusion of multi-exposure low dynamic range (LDR) images captured using a Lytro Illum light field camera. Experimental results show that, with a sampling rate as low as 2.67 %, using two shots, our proposed method yields a higher light field reconstruction quality compared to the fusion of multiple LDR light fields captured with different exposures, and with the fusion of multiple LDR light fields captured with different ISO settings. This framework leads to a new design for single sensor compressive HDR light field cameras, combining multi-ISO photography with coded mask acquisition, placed in a compressive sensing framework.

8.1.3 Depth estimation at the decoder in the MPEG-I standard

Participants: Patrick Garus, Christine Guillemot, Thomas Maugey.

Immersive video often refers to multiple views with texture and scene geometry information, from which different viewports can be synthesized on the client side. To design efficient immersive video coding solutions, it is desirable to minimize bitrate, pixel rate and complexity. We have investigated whether the classical approach of sending the geometry of a scene as depth maps is appropriate to serve this purpose. Previous work has shown that bypassing depth transmission entirely and estimating depth at the client side improves the synthesis performance while saving bitrate and pixel rate. In order to understand if the encoder side depth maps contain information that is beneficial to be transmitted, we have first explored a hybrid approach which enables partial depth map transmission using a block-based RD-based decision in the depth coding process [7]. This approach has revealed that partial depth map transmission may improve the rendering performance but does not present a good compromise in terms of compression efficiency. This led us to address the remaining drawbacks of decoder side depth estimation: complexity and depth map inaccuracy. We propose a novel system that takes advantage of high quality depth maps at the server side by encoding them into lightweight features that support the depth estimator at the client side. These features have allowed reducing the amount of data that has to be handled during decoder side depth estimation by 88%, which significantly speeds up the cost computation and the energy minimization of the depth estimator. Furthermore, -46.0% and -37.9% average synthesis BD-Rate gains are achieved compared to the classical approach with depth maps estimated at the encoder.

8.1.4 A Light Field FDL-HCGH Feature in Scale-Disparity Space

Participants: Christine Guillemot.

Many computer vision applications heavily rely on feature detection, description, and matching. Feature detectors are mainly based on specific image gradient distributions, which have local or global invariance to possible image translation, rotation, or to scale or affine transformation. The identifiability and invariance of features description are critical in feature matching.

In collaboration with Xi'an University (Prof. Zhaolin Xiao), we have proposed novel feature descriptors for light fields computed on the Fourier disparity layer representations of the light field (see Fig.(4)). A first feature extraction taking advantage of both the Harris feature detector and the SIFT descriptor has been proposed in [22]. We have then developed a second feature descriptor, called FDL-HCGH feature, which is based on the Harris detection in a scale-disparity space, and a circular gradient histogram descriptor. It is shown to yield more accurate feature matching, compared with the reference LIght Field Feature (LiFF) descriptor, with a lower computational complexity. In order to evaluate the feature matching performance with the proposed descriptor, we have generated a synthetic stereo LF dataset with ground truth matching points. Experimental results with synthetic and real-world datasets show

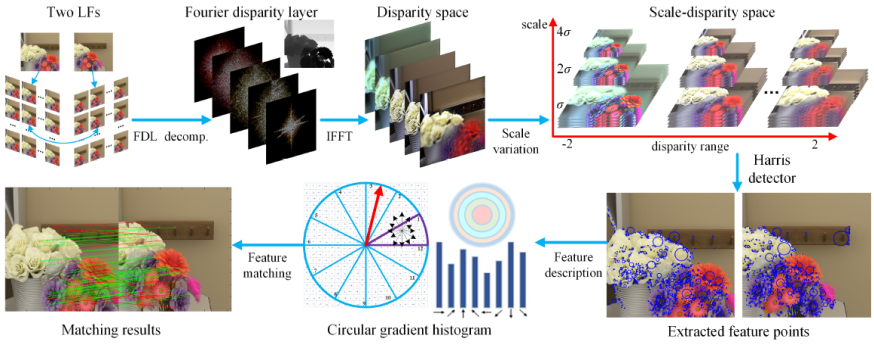


Figure 4: Light field feature computation and matching for a stereo pair of light fields. Harris points are first computed in the scale-disparity space of the two light field Fourier disparity layer representations. The descriptors are then derived by computing circular gradient histograms and used for finding feature point matches between the two input light fields.

that our solution outperforms existing methods in terms of both feature detection robustness and feature matching accuracy.

8.2 Signal processing and learning methods for visual data representation and compression

Keywords: Sparse representation, data dimensionality reduction, compression, scalability, rate-distortion theory.

8.2.1 Plenoptic point cloud compression

Participants: Christine Guillemot, Maja Krivokuca.

In collaboration with Google (Phil Chou) and the Univ. of Linkoping (Ehsan Miandji), we have introduced a novel 6-D representation of plenoptic point clouds, enabling joint, non-separable transform coding of plenoptic signals defined along both spatial and angular (viewpoint) dimensions [9]. This 6-D

representation, which is built in a global coordinate system, can be used in both multi-camera studio capture and video fly-by capture scenarios, with various viewpoint (camera) arrangements and densities. We show that both the Region-Adaptive Hierarchical Transform (RAHT) and the Graph Fourier Transform (GFT) can be extended to the proposed 6-D representation to enable the non-separable transform coding. Our method is applicable to plenoptic data with either dense or sparse sets of viewpoints, and to complete or incomplete plenoptic data, while the state-of-the-art RAHT-KLT method, which is separable in spatial and angular dimensions, is applicable only to complete plenoptic data. The "complete" plenoptic data refers to data that has, for each spatial point, one color for every viewpoint (ignoring any occlusions), while "incomplete" data has colors only for the visible surface points at each viewpoint. We have demonstrated that the proposed 6-D RAHT and 6-D GFT compression methods are able to outperform the state-of-the-art RAHT-KLT method on 3-D objects with various levels of surface specularity, and captured with different camera arrangements and different degrees of viewpoint sparsity.

8.2.2 Graph coarsening and dimensionality reduction for graph transforms of reduced complexity

Participants: Christine Guillemot, Thomas Maugey.

Graph-based transforms are powerful tools for signal representation and energy compaction. However, their use for high dimensional signals such as light fields poses obvious problems of complexity. To overcome this difficulty, one can consider local graph transforms defined on supports of limited dimension, which may however not allow us to fully exploit long-term signal correlation. We have developed methods to optimize local graph supports in a rate distortion sense for efficient light field compression [13]. A large graph support can be well adapted for compression efficiency, however at the expense of high complexity. In this case, we use graph reduction techniques to make the graph transform feasible. We also considered spectral clustering to reduce the dimension of the graph supports while controlling both rate and complexity (see Fig.(5)) for an example of segmentation resulting from spectral clustering). We derived the distortion and rate models which are then used to guide the graph optimization. We developed a complete light field coding scheme based on the proposed graph optimization tools. Experimental results show rate-distortion performance gains compared to the use of fixed graph support. The method also provides competitive results when compared against HEVC-based and the JPEG Pleno light field coding schemes. WE also assess the method against a homography-based low rank approximation and a Fourier disparity layer based coding method.

Figure 5: An example of obtained segmentation map.

8.2.3 Untrained Neural Network Prior for Light Field Representation and Compression

Participants: Christine Guillemot, Xiaoran Jiang, Jinglei Shi.

Deep generative models have proven to be effective priors for solving a variety of image processing problems. However, the learning of realistic image priors, based on a large number of parameters, requires a large amount of training data. It has been shown recently, with the so-called deep image prior (DIP), that randomly initialized neural networks can act as good image priors without learning.

We have proposed a deep generative model for light fields, which is compact and which does not require any training data other than the light field itself. The proposed network is based on both a generative model that aims at modeling the spatial information that is static, i.e., found in all light field views, and on a convolutional Gated Recurrent Unit (ConvGRU) that is used to model variations between angular views. The spatial view generative model is inspired from the deep decoder, itself built upon the deep image prior, but that we enhance with spatial and channel attention modules, and with quantization-aware learning. The attention modules modulate the feature maps at the output of the different layers of the generator. In addition, we offer an option which expressively encodes the upscaling operations in learned weights in order to better fit the light field to process. The deep decoder is also adapted in order to model several light field views, with layers (i.e. features) that are common to all views and others that are specific to each view. The weights of both the convGRU and the deep decoder are learned end-to-end in order to minimize the reconstruction error of the target light field.

To show the potential of the proposed generative model, we have developed a complete light field compression scheme with quantization-aware learning and entropy coding of the quantized weights. Experimental results show that the proposed method outperforms state-of-the-art light field compression methods as well as recent deep video compression methods in terms of both PSNR and MS-SSIM metrics.

8.2.4 OSLO: On-the-Sphere Learning for Omnidirectional images and its application to 360-degree image compression

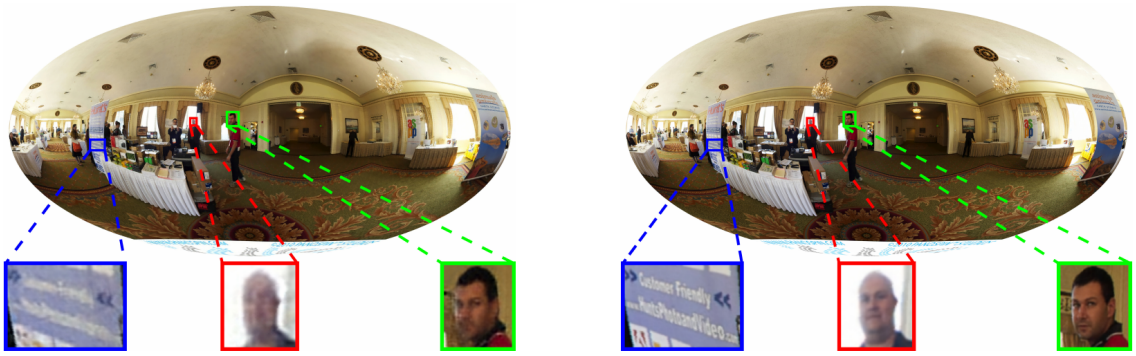
Participants: Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy.

State-of-the-art 2D image compression schemes rely on the power of convolutional neural networks (CNNs). Although CNNs offer promising perspectives for 2D image compression, extending such models to omnidirectional images is not straightforward. First, omnidirectional images, when compressed on 2D maps, have specific spatial and statistical properties that can not be fully captured by current CNN models. Second, basic mathematical operations composing a CNN architecture, e.g., translation and sampling, are not well-defined when the omnidirectional image is defined directly on the sphere. We have studied the learning of representation models for on-the-sphere omnidirectional images and we have proposed to use the properties of HEALPix uniform sampling of the sphere to redefine the mathematical tools used in deep learning models for omnidirectional images. In particular, i) we have proposed the definition of a new convolution operation on the sphere that keeps the high expressiveness and the low complexity of a classical 2D convolution; ii) we have adapted standard CNN techniques such as stride, iterative aggregation, and pixel shuffling to the spherical domain; and then iii) we have applies our new framework to the task of omnidirectional image compression. Our experiments shown that our proposed on-the-sphere solution leads to a better compression gain that can save 13.7% of the bit rate compared to similar learned models applied to equirectangular images. Also, compared to learning models based on graph convolutional networks, our solution supports more expressive filters that can preserve high frequencies and provide a better perceptual quality of the compressed images. Such results demonstrate the efficiency of the proposed framework, which opens new research venues for other omnidirectional vision tasks to be effectively implemented on the sphere manifold.

8.2.5 Rate-distortion optimized motion estimation for on-the-sphere compression of 360 videos

Participants: Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy.

On-the-sphere compression of omnidirectional videos is a very promising approach. First, it saves computational complexity as it avoids to project the sphere onto a 2D map, as classically done. Second,



(a) 2408.43 KB (b) 936.65 KB

Figure 6: Visual comparison of decoded images when compressed with (a) DeepSphere (the competitor) and (b) our OSLO solution.

and more importantly, it allows to achieve a better rate-distortion tradeoff, since neither the visual data nor its domain of definition are distorted. In [20], the on-the-sphere compression for omnidirectional still images, previously developed, is extended to videos. We have first proposed a complete review of existing spherical motion models. Then we have proposed a new one called tangent-linear+t. We have finally proposed a rate-distortion optimized algorithm to locally choose the best motion model for efficient motion estimation/compensation. For that purpose, we have additionally proposed a finer search pattern, called spherical-uniform, for the motion parameters, which leads to a more accurate block prediction. The novel algorithm leads to rate-distortion gains compared to methods based on a unique motion model.

8.2.6 Satellite image compression and restoration

Participants: Denis Bacchus, Christine Guillemot, Arthur Lecert, Aline Roumy.

In the context of the Lichie project, in collaboration with Airbus, we address two problems for satellite imaging: quasi-lossless compression and restoration, using deep learning methods.

More precisely, we developed an end-to-end trainable neural network for satellite image compression. The proposed approach builds upon an image compression scheme based on variable autoencoders with a learned hyper-prior that captures dependencies in the latent space for entropy coding. We explore this architecture in light of specificities of satellite imaging: processing constraints on board the satellite (complexity and memory constraints) and quality needed in terms of reconstruction for the processing task on the ground. We explored data augmentation to improve the reconstruction of challenging image patterns. The proposed model outperforms the current standard of lossy image compression onboard satellite-based on JPEG 2000, as well as the initial hyperprior architecture designed for natural images.

In parallel, we have developed a method to estimate the components of the Retonex model using untrained deep generative networks to restore low light satellite images. The Retinex model has indeed been shown to be an effective tool for low-light image restoration. This model assumes that an image can be decomposed into a product of two components, the illumination and the reflectance. Efficient methods have been proposed to estimate these components based on deep neural networks trained in a supervised manner with a dataset of paired low/normal-light images. However, collecting these samples is extremely challenging in practice. The proposed approach does not require any training data other than the input low-light image. To demonstrate the efficiency of the proposed estimation method, we perform simple gamma corrections on the illumination and reflectance components. We show that our approach leads to better restoration results than existing unsupervised methods and on par with fully supervised solutions thanks to the decomposition process.

8.2.7 Neural networks for video compression acceleration

Participants: Christine Guillemot, Yiqun Liu, Aline Roumy.

In the context of the Cifre contract with Ateme, we investigate deep learning architectures for the inference of coding modes in video compression algorithms with the ultimate goal of reducing the encoder complexity. In particular, we studied the recently finalized video compression standard VVC. Compared to its predecessor standard HEVC, VVC offers about 50% compression efficiency gain, in terms of rate, at the cost of about 10x more encoder complexity. We therefore constructed a CNN-based method to speed up the partitioning of an image into blocks. More precisely, an image is first split into fixed-size so called coding tree unit. Then, each CTU is partitioned into blocks called CU which are adapted to the content. This operation, being adapted to the content, is of a extreme computational complexity, as it requires to perform for each possible partition, the whole encoding and its Rate-distortion optimization. The proposed CNN allows to avoid to test partitions that are unlikely to be selected. Thanks to a light-weight CNN, experiments show that the proposed method can achieve acceleration ranging from 17% to 35% with a reasonable efficiency drop ranging from 0.32% to 1.21% in terms of rate.

8.2.8 Multiple profile video compression optimization

Participants: Reda Kaafarani, Thomas Maugey, Aline Roumy.

In the context of the Cifre contract with MediaKind, we develop coding tools in order to compress and deliver video, while adapting the quality to the available bandwidth and/or the user screen resolution. As a first step towards this goal, we studied in [19] bitrate ladders for the last standardized video coder, named Versatile Video Coder (VVC). Indeed, many video service providers take advantage of bitrate ladders in adaptive HTTP video streaming to account for different network states and user display specifications by providing bitrate/resolution pairs that best fit client's network conditions and display capabilities. These bitrate ladders, however, differ when using different codecs and thus the couples bitrate/resolution differ as well. In addition, bitrate ladders are based on previously available codecs (H.264/MPEG4-AVC, HEVC, etc.), i.e. codecs that are already in service, hence the introduction of new codecs e.g. VVC requires re-analyzing these ladders. For that matter, we analyzed the evolution of the bitrate ladder when using VVC. We showed how VVC impacts this ladder when compared to HEVC and H.264/AVC and in particular, that there is no need to switch to lower resolutions at the lower bitrates defined in the Call for Evidence on Transcoding for Network Distributed Video Coding (CfE).

8.3 Algorithms for inverse problems in visual data processing

Keywords: Inpainting, denoising, view synthesis, super-resolution.

8.3.1 Deep light field view synthesis and temporal interpolation

Participants: Simon Evain, Christine Guillemot, Xiaoran Jiang, Jinglei Shi.

We have pursued our development of a learning-based framework for light field view synthesis from a subset of input views, for which we published preliminary results at CVPR 2020. We have in particular proposed a deep residual architecture that can be used both for synthesizing high quality angular views in light fields and temporal frames in classical videos. The proposed framework consists of an optical flow estimator optimized for view synthesis, a trainable feature extractor and a residual convolutional network for pixel and feature-based view reconstruction. Among these modules, the fine-tuning of the optical flow estimator specifically for the view synthesis task yields scene depth or motion information that is well optimized for the targeted problem. In cooperation with the end-to-end trainable encoder, the synthesis

block employs both pixel-based and feature-based synthesis with residual connection blocks, and the two synthesized views are fused with the help of a learned soft mask to obtain the final reconstructed view. Experimental results with various datasets show that our method performs favorably against other state-of-the-art methods with a large gain for light field view synthesis. Furthermore, with a little modification, our method can also be used for video frame interpolation, generating high quality frames compared with existing interpolation methods. We have also proposed a specific deep learning-based network for video frame rate up-conversion (or video frame interpolation)in [21]. The proposed optical flow-based pipeline employs deep features extracted to learn residue maps for progressively refining the synthesized intermediate frames [24].

We have also proposed a learning-based method to extrapolate novel views from axial volumes of sheared epipolar plane images (EPIs), which allows us to increase the axial light field resolution [14]. Axial light field resolution refers to the ability to distinguish features at different depths by refocusing. With the proposed method, the extrapolated light field gives re-focused images with a shallower depth of field (DOF), leading to more accurate refocusing results. The refocusing precision can be essential for some light field applications like microscopy. The proposed approach does not need accurate depth estimation. Experimental results with both synthetic and real light fields show that the method not only works well for light fields with small baselines as those captured by plenoptic cameras, but also applies to light fields with larger baselines.

Finally, we have designed s lightweight neural network architecture with an adversarial loss for generating a full light field from one single image [17]. The method is able to estimate disparity maps and automatically identify occluded regions from one single image thanks to a disparity confidence map based on forward-backward consistency checks. The disparity confidence map also controls the use of an adversarial loss for occlusion handling. The approach outperforms reference methods when trained and tested on light field data. Besides,we also designed the method so that it can efficiently generate a full light field from one single image, even when trained only on stereo data. This allows us to generalize our approach for view synthesis to more diverse data and semantics [23].

8.3.2 Optimization methods with learned priors

Participants: Rita Fermanian, Christine Guillemot, Brandon Le Bon, Mikael Le Pendu.

Recent methods have been introduced with the goal of combining the advantages of well understood iterative optimization techniques with those of learnable complex image priors. A first category of methods, referred to as "Plug-and-play" methods, has been introduced where a learned network-based prior is plugged in an iterative optimization algorithm. These learnable priors can take several forms, the most common ones being: a projection operator on a learned image subspace, a proximal operator of a regularizer or a denoiser.

In the context of the AI chair DeepCIM, we have first studied Plug-and-Play optimization for solving inverse problems by plugging a denoiser into a classical optimization algorithm. The denoiser accounts for the regularization and therefore implicitly determines the prior knowledge on the data, hence replacing typical handcrafted priors. We have extended the concept of plug-and-play optimization to use denoisers that can be parameterized for non-constant noise variance. In that aim, we have introduced a preconditioning of the ADMM algorithm, which mathematically justifies the use of such an adjustable denoiser. We additionally proposed a procedure for training a convolutional neural network for high quality non-blind image denoising that also allows for pixel-wise control of the noise standard deviation. We have shown that our pixel-wise adjustable denoiser, along with a suitable preconditioning strategy, can further improve the plug-and-play ADMM approach for several applications, including image completion, interpolation, demosaicing and Poisson denoising. An illustration of Poisson denoising results is given in Fig.(7).

One advantage of such learned priors is their genericity in the sense that they can be used for any inverse problem, and do not need to be re-trained for each new problem, in contrast with deep models learned as a regression function for a specific task. However, priors learned independently of the targeted problem may not yield the best solution. Unrolling a fixed number of iterations of optimization algorithms

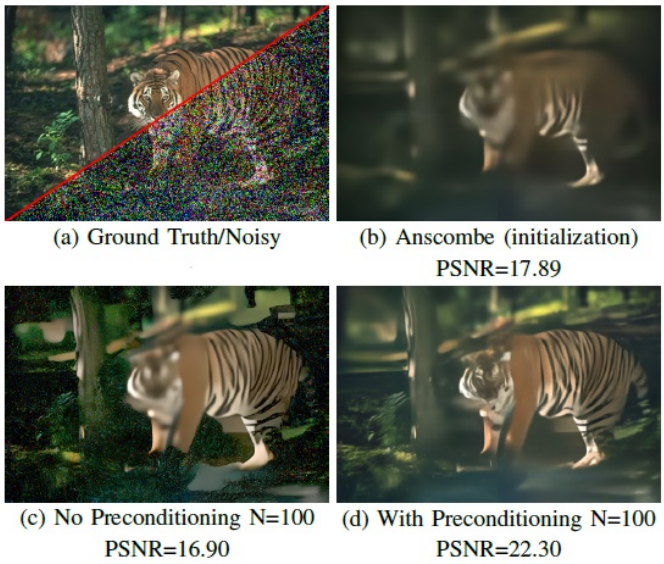


Figure 7: Poisson Denoising for very high noise level. N is the number of iterations of the ADMM iterative algorithms.

is another way of coupling optimization and deep learning techniques. The learnable network is trained end to- end within the iterative algorithm so that performing a fixed number of iterations yields optimized results for a given inverse problem. Several optimization algorithms (Iterative Shrinkage Thresholding Algorithm (ISTA), Half S Quadratic Splitting (HQS), and Alternating Direction Method of Multipliers (ADMM)) have been unrolled in the literature, where a learned regularization network is used at each iteration of the optimization algorithm.

While usual iterative methods iterate until idempotence, i.e. until the difference between the input and the output is sufficiently small, the number of iterations in unrolled optimization methods is set to a small value. This makes it possible to learn a component end-to-end within the optimization algorithm, hence in a way which takes into account the data term, i.e., the degradation operator. But learning networks end-to-end within an unrolled optimization scheme requires high GPU memory usage since the memory used for the backpropagation scales linearly with the number of iterations. This explains why the number of iterations used in an unrolled optimization method is limited. To cope with these limitations, we have developed a stochastic implicit unrolled proximal point algorithm with a learned denoiser, in which sub-problems are defined per iteration. We exploit the fact that the Douglas-Rachford algorithm is an application of the proximal point algorithm to re-define the unrolled step as a proximal mapping. We focused on the unrolled ADMM, which has been demonstrated to be a special case of the Douglas-Rachford algorithm, hence of the proximal point algorithm. This allows us to introduce a novel unrolled proximal gradient method coupling an implicit model and a stochastic learning strategy. We have shown that this stochastic iteration update strategy better controls the learning at each unrolled optimization step, hence leads to a faster convergence than other implicit unrolled methods, while maintaining the advantage of a low GPU memory usage, as well as similar reconstruction quality to the best unrolled methods for all considered image inverse problems.

We have also considered untrained generative model and proposed an optimization method coupling a learned denoiser with the untrained generative model, called deep image prior (DIP) in the framework of the Alternating Direction Method of Multipliers (ADMM) method [18]. We have also studied different regularizers of DIP optimization, for inverse problems in imaging, focusing in particular on denoising and super-resolution. The goal was to make the best of the untrained DIP and of a generic regularizer learned in a supervised manner from a large collection of images. When placed in the ADMM framework, the denoiser is used as a proximal operator and can be learned independently of the considered inverse problem. We show the benefits of the proposed method, in comparison with other regularized DIP methods, for two linear inverse problems, i.e., denoising and super-resolution

8.4 User centric compression

Keywords: Information theory, interactive communication, coding for machines, generative compression, database sampling.

8.4.1 Interactive compression

Participants: Sebastien Bellenous, Nicolas Charpenay, Thomas Maugey, Aline Roumy.

In the Intercom project, we have studied the impact of interactivity on the coding performance. We have, for example, tackled the following problem: is it possible to compress a 360-degree video once for all, and then partly extract and decode what is needed for a user navigation, while, keeping good compression performance? First, we derived the achievable theoretical bounds in terms of storage and transmission rates. In [15], we analyzed and improved a practical coding scheme. We considered a binarized coding scheme, which insures a low decoding complexity. First, we showed that binarization does not impact the transmission rate but only slightly the storage with respect to a symbol based approach. Second, we proposed a Q-ary symmetric model to represent the pairwise joint distribution of the sources instead of the widely used Laplacian model. Third, we introduced a novel pre-estimation strategy, which allows to infer the symbols of some bit planes without any additional data and therefore permits to reduce the storage and transmission rates. In the context of 360° images, the proposed scheme allows us to save 14% and 34% bitrate in storage and transmission rates respectively.

Previously, we derived information theoretical bounds of the compression problem with interactivity under a vanishing error probability assumption, a classical framework in information theory. In practical systems however, the core algorithm (entropy coder) needs to achieve exactly zero-error for any blocklength. Therefore, to complete our previous work, it was of great importance to also derive the compression performance in a zero-error framework. To do so, we modeled in [16] the interactive compression problem as a source coding problem when side-information (SI) may be present. Indeed, the side information may represent an image that could have been requested previously by the user. In particular, we showed that both zero-error and vanishing error schemes achieve exactly the same asymptotic compression performance. The proof technique relied on a random coding argument, and a code construction based on coset partitioning obtained from a linear code.

In the Intercom project, after deriving the achievable compression performance, we have built a new omnidirectional video coder. This original architecture enables to reduce significantly the cost of interactivity compared to the conventional video coders. In the project ICOV, we are developing a complete and well specified coder that is aimed to be shared with the community, starting from this promising proof of concept. In the year 2021, we have first worked on the bitstream specification. As for the video standards, we have defined the structure of the binary code that is stored on the servers and transmitted to the decoder. Then, we have worked on the implementation of the hierarchical channel coder that is the new entropy coding strategy that enables flexible decoding. We have tested the performance and compared them to the theoretical Shannon entropy, demonstrating the small gap between the theoretical results and the one achieved in practice.

8.4.2 Data Repurposing

Participants: Anju Jose Tom, Tom Bachard, Thomas Maugey.

Compression algorithms are nowadays overwhelmed by the tsunami of visual data created everyday. Despite a growing efficiency, they are always constrained to minimize the compression error, computed in the pixel domain.

The *Data Repurposing*, proposed in the team, reinvents how compression is done. It consists in semantically describing the database information in a concise representation, thus leading to drastic

compression ratios *exactly as a music score is able to describe, for example, a concert in a compact and reusable form.* This enables the compression to withdraw tremendous amount of useless, or at least not essential, information while condensing the important information into a compact recycled signal. In a nutshell, in the Data Repurposing framework, the decoded signals target subjective exhaustiveness of the information description, rather than fidelity to the input data, as in the traditional compression algorithms. In the exploratory action DARE, we had the chance to explore two directions.

We first introduce the concept of perceived information (PI), which reflects the information perceived by a given user experiencing a data collection, and which is evaluated as the volume spanned by the sources features in a *personalized* latent space. We use this PI metric in order to formalize a database sampling algorithm. The goal is to take into account the user's preferences while keeping a certain level of diversity in the sampled database (Figure 8). A first version of our algorithm outperforms benchmark solutions with simulation results, showing the gain in taking into account users' preferences while also maximizing the perceived information in the feature domain. We are currently working on the extension of such algorithm for real images as inputs.

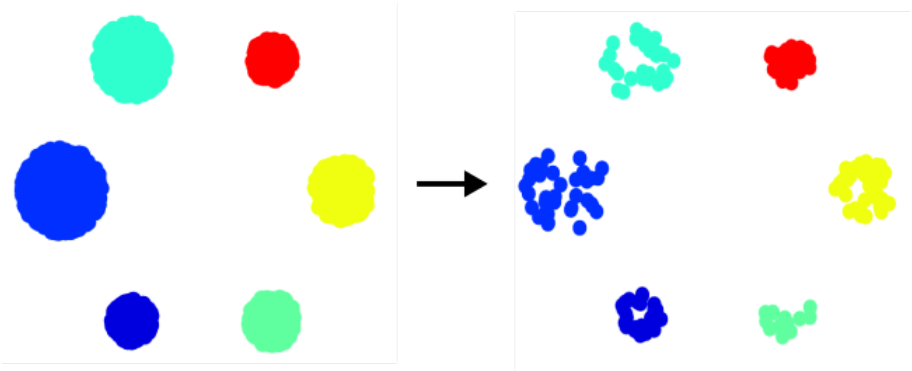


Figure 8: Database sampling. Each point corresponds to an item of the data collection. The color indicates the proximity in terms of semantic.

A second direction for Data Repurposing explored this year deals with the generative compression. Basically, it consists in allowing the compression algorithm to "reinvent" part of the data at the decoding phase, and thus saving a lot of bit-rate by not coding it. This work is currently at a preliminary stage. We have started our study on the possibility of shaping the latent space of the compressed description such that it includes a description of the semantic. In 2021, a Young researcher ANR project has also been accepted on this research theme. It will begin in April 2022 officially.

9 Bilateral contracts and grants with industry

9.1 Bilateral contracts with industry

CIFRE contract with Orange labs. on compression of immersive content

Participants: Patrick Garus, Christine Guillemot, Thomas Maugey.

- Title : Compression of immersive content
- Research axis : 8.1.3
- Partners : Orange labs. (F. Henry), Inria-Rennes.
- Funding : Orange, ANRT.
- Period : Jan.2019-Dec.2021.

The goal of this Cifre contract is to develop novel compression methods for 6 DoF immersive video content. This implies investigating depth estimation and view synthesis methods that would be robust to quantization noise, for which deep learning solutions are being considered. This also implies developing the corresponding coding mode decisions based on rate-distortion criteria.

CIFRE contract with Ateme on neural networks for video compression

Participants: Christine Guillemot, Yiqun Liu, Aline Roumy.

- Title : Neural networks for video compression of reduced complexity
- Partners : Ateme (T. Guionnet, M. Abdoli), Inria-Rennes.
- Funding : Ateme, ANRT.
- Period : Aug.2020-Jul.2023.

The goal of this Cifre contract is to investigate deep learning architectures for the inference of coding modes in video compression algorithms with the ultimate goal of reducing the encoder complexity. The first step addresses the problem of Intra coding modes and quad-tree partitioning inference. The next step will consider Inter coding modes taking into account motion and temporal information.

Contract LITCHIE with Airbus on deep learning for satellite imaging

Participants: Denis Bacchus, Christine Guillemot, Arthur Lecert, Aline Roumy.

- Title : Deep learning methods for low light vision
- Partners : Airbus (R. Fraisse), Inria-Rennes.
- Funding : BPI.
- Period : Sept.2020-Aug.2023.

The goal of this contract is to investigate deep learning methods for low light vision with sattelite imaging. The SIROCCO team focuses on two complementary problems: compression of low light images and restoration under conditions of low illumination, and hazing. The problem of low light image enhancement implies handling various factors simultaneously including brightness, contrast, artifacts and noise. We investigate solutions coupling the retinex theory, assuming that observed images can be decomposed into reflectance and illumination, with machine learning methods. We address the compression problem taking into account the processing tasks considered on the ground such as the restoration task, leading to an end-to-end optimization approach.

Research collaboration contract with MediaKind on Video encoding optimization for the Versatile Video Coding standard VVC

Participants: Reda Kaafarani, Thomas Maugey, Aline Roumy.

- Title : Multiple profile encoding optimization
- Partners : MediaKind, Inria-Rennes.
- Funding : MediaKind.

- Period : Dec.2020-April 2021.

The goal of this study is to analyze the video compression standard recently standardized and called Versatile Video Coding (VVC) in the context of streaming. In particular, the ultimate goal is to provide the users with the best user experience in other words the best tradeoff between rate and complexity taking into account the bandwidth limitation at the user side. This optimization is performed while adjusting both the resolution of the video and the quantization level, and the optimization result is given in terms of a curve called bitrate-ladder and provides for each user bandwidth rate, the best video encoder configuration (resolution quantization).

Cifre contract with MediaKind on Multiple profile encoding optimization

Participants: Reda Kaafarani, Thomas Maugey, Aline Roumy.

- Title : Multiple profile encoding optimization
- Partners : MediaKind, Inria-Rennes.
- Funding : MediaKind, ANRT.
- Period : April 2021-April 2024.

The goal of this Cifre contract is to optimize a streaming solution taking into the whole process, namely the encoding, the long-term and the short term storages (in particular for replay, taking into the popularity of the videos), the multiple copies of a video (to adapt to both the resolution and the bandwidth of the user), and the transmissions (between all entities: encoder, back-end and front-end server, and the user). This optimization will be with several objectives as well. In particular, the goals will be to maximize the user experience but also to save energy and/or the deployment cost of a streaming solution.

10 Partnerships and cooperations

10.1 European initiatives

10.1.1 FP7 & H2020 projects

ERC-CLIM: Computational Light Field Imaging

Participants: Simon Evain, Christine Guillemot, Xiaoran Jiang, Guillaume Le Guludec, Hoai Nam Nguyen, Jinglei Shi.

- Partners : Inria-Rennes,
- Funding : European Commission
- Period : Sept.2016-Feb.2022.

All imaging systems, when capturing a view, record different combinations of light rays emitted by the environment. In a conventional camera, each sensor element sums all the light rays emitted by one point over the lens aperture. Light field cameras instead measure the light along each ray reaching the camera sensors and not only the sum of rays striking each point in the image. In one single exposure, they capture the geometric distribution of light passing through the lens. This process can be seen as sampling the plenoptic function that describes the intensity of the light rays interacting with the scene and received by an observer at every point in space, along any direction of gaze, for all times and every wavelength.

The recorded flow of rays (the light field) is in the formof high-dimensional data (4D or 5D for static and dynamic light fields). The 4D/5D light field yields a very rich description of the scene enabling advanced creation of novel images from a single capture, e.g. for computational photography by simulating a capture with a different focus and a different depth of field, by simulating lenses with different apertures, by creating images with different artistic intents. It also enables advanced scene analysis with depth and scene flow estimation and 3D modeling. The goal of the ERC-CLIM project is to develop algorithms for the entire static and video light fields processing chain. The planned research includes the development of:

- Acquisition methods and novel coded-mask based camera models,
- Novel low-rank or graph-based models for dimensionality reduction and compression
- Deep learning methods for scene analysis (e.g. scene depth and scene flow estimation)
- Learning methods for solving a range of inverse problems: denoising, super-resolution, axial super-resolution, view synthesis.

H2020 Marie Sklodowska-Curie Innovative Training Network Plenoptima: plenoptic Imaging

Participants: Ipek Anil Atalay, Davi Freitas, Kai Gu, Christine Guillemot, Soheib Takhtardeshir, Samuel Willingham.

- Partners : MidSweden Univ., Tampere Univ., Technical Univ. Berlin, Inria-Rennes, Institute of Optical Materials and Technologies, Bulgarian Academy of Sciences.
- Funding : European Commission
- Period : Jan.2021-Dec.2024.

Plenoptic Imaging (PLENOPTIMA) is a four-year (2021–2024) H2020 Marie Sklodowska-Curie Innovative Training Network that develops a cross-disciplinary approach to plenoptic imaging, which includes new optical materials and sensing principles, signal processing methods, new computing architectures, and vision science modelling. The ultimate goal of PLENOPTIMA is to establish new cross-sectorial, international, multi-university sustainable doctoral degree programmes in the area of plenoptic imaging and to train fifteen next generation researchers and creative professionals within these programmes for the benefit of a variety of application sectors.

10.2 National initiatives

10.2.1 Project Action Exploratoire "Data Repurposing"

Participants: Anju Jose Tom, Thomas Maugey.

- Funding: Inria.
- Period: Sept. 2020 - Aug. 2023.

Lossy compression algorithms trade bits for quality, aiming at reducing as much as possible the bitrate needed to represent the original source (or set of sources), while preserving the source quality. In the exploratory action "DARE", we propose a novel paradigm of compression algorithms, aimed at minimizing the *information loss perceived by the final user* instead of the actual source quality loss, under compression rate constraints. In particular, we plan to measure the amount of information spanned by a data collection in the semantic domain. First, it enables to identify the high-level information contained in each of the image/video of a data collection. Second, it permits to take into account the redundancies

and dissmilarities in the calculation of the global volume of information that is contained in a data collection. Finally, we propose to take into account the user's preferences in this calculation, since two users may have different tastes and priorities. Once the measure of information is set, we plan to build efficient sampling algorithms to reduce the data collection's size. This project also enables to explore new ideas for image generative compression, when part of the content can be "invented" at the decoder side.

10.2.2 IA Chair: DeepCIM- Deep learning for computational imaging with emerging image modalities

Participants: Rita Fermanian, Christine Guillemot, Brandon Lebon, Guillaume Le Guludec, Mikael Le Pendu, Jinglei Shi.

- Funding: ANR (Agence Nationale de la Recherche).
- Period: Sept. 2020 - Aug. 2024.

The project aims at leveraging recent advances in three fields: image processing, computer vision and machine (deep) learning. It will focus on the design of models and algorithms for data dimensionality reduction and inverse problems with emerging image modalities. The first research challenge will concern the design of learning methods for data representation and dimensionality reduction. These methods encompass the learning of sparse and low rank models, of signal priors or representations in latent spaces of reduced dimensions. This also includes the learning of efficient and, if possible, lightweight architectures for data recovery from the representations of reduced dimension. Modeling joint distributions of pixels constituting a natural image is also a fundamental requirement for a variety of processing tasks. This is one of the major challenges in generative image modeling, field conquered in recent years by deep learning. Based on the above models, our goal is also to develop algorithms for solving a number of inverse problems with novel imaging modalities. Solving inverse problems to retrieve a good representation of the scene from the captured data requires prior knowledge on the structure of the image space. Deep learning based techniques designed to learn signal priors, tcan be used as regularization models.

10.2.3 CominLabs MOVE project: Mature Omnidirectional Video Exploration.

Participants: Sébastien Bellenous, Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy.

- Funding: Labex CominLabs.
- Period: Jan. 2021 - Dec. 2022.

This project aims to secure the industrial impact of the InterCom project, ended in Dec. 2020, and funded by the Labex CominLabs. Indeed, the goal of the Cominlabs InterCom project was to design novel compression algorithms to allow interactive communication between users and a server. One instance of this interactive scenario is a visual immersive experience, where a user can navigate freely in a 3D scene. This leads to tremendous amount of visual data, such that the whole scene cannot be sent to the user. Fortunately, the user watches only a part of the scene, such that only the requested part needs to be sent to the user. This however presents a great challenge from the point of view of the compression. Indeed, the data needs to be compressed once online, but decompressed in many manners, one manner being one requested point of view. In the case of a static 360 camera, the navigation has 3 Degrees of Freedom such that there is one decompression per value of a 3D-vector. One of the achievement of the InterCom project is a prototype for an interactive compression scheme for 360-degree images. The MOVE project helps the ADT ICOV project towards the construction of a full demonstrator. It also helps the maturation of the startup project lead by Navid Mahmoudian Bidgoli (ANAX).

10.2.4 CominLabs Colearn project: Coding for Learning

Participants: Thomas Maugey, Rémi Piau, Aline Roumy.

- Partners: Inria-Rennes (Sirocco team); LabSTICC, IMT Atlantique, (team Code and SI); IETR, INSA Rennes (Syscom team).
- Funding: Labex CominLabs.
- Period: Sept. 2021 - Dec. 2024.

The amount of data available online is growing so fast that it is essential to rely on advanced Machine Learning techniques so as to automatically analyze, sort, and organize the content uploaded by e.g. sensors or users. The conventional data transmission framework assumes that the data should be completely reconstructed, even with some distortions, by the server. Instead, this project aims to develop a novel communication framework in which the server may also apply a learning task over the coded data. The project will therefore develop an Information Theoretic analysis so as to understand the fundamental limits of such systems, and develop novel coding techniques allowing for both learning and data reconstruction from the coded data.

10.2.5 Inria Start-up Studio: Anax

Participants: Navid Mahmoudian Bidgoli, Simon Evain, Thomas Maugey, Aline Roumy.

Two former PhD students of the team (Navid Mahmoudian Bidgoli and Simon Evain) plan to launch a startup, named Anax, on the theme of omnidirectional/image processing. They have obtained a one year grant (Sept 2021 - Aug 2022) for both of them funded by the Inria Startup Studio. Here is the description of the Anax project.

Anax aims to provide a deep tech software solution for processing 360-degree visual content with artificial intelligence (AI) specially designed for the preparation of virtual tours. Anax is aimed at various actors who wish to offer an immersive visit experience to improve their visibility, such as real estate agencies, cultural institutions, and interior designers. Anax is developing a technology that allows retrieving a faithful 3D reconstruction of a building from 360-degree images, opening up a wide range of applications based on AI image processing such as automatic recommendation of similar apartments, augmented reality, automatic inventory, etc. The envisaged solution, based on artificial intelligence, works even with consumer 360-degree capturing devices that are readily accessible to the general public.

11 Dissemination

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

General chair, scientific chair

- C. Guillemot has organized the ERC-CLIM workshop, 29-30th Sept. 2021.
- A. Roumy and T. Maugey have organized a thematic day for the GdR-Isis on “Learning Based Coding for Digital Image and Video Information”, 15th June 2021

Member of the organizing committees

- C. Guillemot was a member of the organizing committee of the Picture Coding Symposium, Bristol, 29th June-2nd July. 2021
- C. Guillemot was a member of the organizing committee (keynote chair) of the IEEE Multimedia Workshop, Tampere, 6-8 Oct. 2021
- T. Maugey was a member of the organizing committee (awards chair) of the IEEE Multimedia Workshop, Tampere, 6-8 Oct. 2021

11.1.2 Scientific events: selection

Member of the conference program committees

- A. Roumy was a member of the technical program committee of the CVPR 2021 workshop on New Trends in Image Restoration and Enhancement (NTIRE).

11.1.3 Journal

Member of the editorial boards

- C. Guillemot is Associate Editor of the IEEE signal processing magazine.
- A. Roumy is Associate Editor of the IEEE Trans. on Image Processing.
- A. Roumy is associate editor of the Springer Annals of Telecommunications.
- T Maugey is Associate Editor of the EURASIP Journal of Advances in Signal Processing

11.1.4 Invited talks

- C. Guillemot gave an invited talk on Light–Field Computational Imaging: Acquisition and Inverse Problems at IMT-Atlantique, March 2021.
- C. Guillemot gave an invited talk on deep compressive light field acquisition at MidSweden University, May 2021.
- T. Maugey gave a seminar at ENS Paris-Saclay on “Compression of visual data: beyond conventional approaches”.

11.1.5 Leadership within the scientific community

- C. Guillemot was chair of the jury of recruitment of CR/ISFP of the Inria center of Nancy Grand Est.
- A. Roumy is member of the IEEE IVMSP technical committee.
- A. Roumy is a Local Liaison Officer for the European Association for Signal Processing (EURASIP).
- A. Roumy is a member of the Executive board of the National Research group in Image and Signal Processing (GRETSI).

11.1.6 Scientific expertise

- C. Guillemot is a member of the Advisory Board of the Strategic Research Programme of the Vrije Universiteit Brussels
- C. Guillemot is responsible of the theme « compression et protection des données images » for the encyclopedia SCIENCES of the publisher ISTE/Wiley (2019-2021).
- C. Guillemot was member of the jury for the Grand Prix Inria, 30 June 2021.

- T. Maugey is a referee for the Italian National Agency for the Evaluation of the University and Research Systems (ANVUR), and for the Digicosme Labex in Paris-Sacaly (Digital worlds: distributed data, programs and architectures).
- A. Roumy was a reviewer for the ANR, for the ANRT, and for the Digicosme Labex in Paris-Sacaly (Digital worlds: distributed data, programs and architectures).

11.1.7 Research administration

- C. Guillemot is a member of the bureau du comité des projets of the center of Rennes Bretagne Atlantique.
- A. Roumy is a member of the Inria evaluation committee.

11.2 Teaching, Supervision, Juries

11.2.1 Teaching

- Master: C. Guillemot, courses on compression sensing and inverse problems, 6 hours, Midsweden Univ., Sweden.
- Master: T. Maugey, course on 3D models in a module on advanced video, 8 hours, M2 SISEA, Univ. of Rennes 1, France.
- Master: T. Maugey, course on Image compression, 10 hours, M2 SISEA, Univ. of Rennes 1, France.
- Master: T. Maugey, course on Representation, editing and perception of digital images, 20 hours, M2 SIF, Univ. of Rennes 1, France.
- Master: T. Maugey, course on 360 video compression, 2.5 hours, M1-M2, IMT Atlantique, France.
- Engineering degree: C. Petit, Sparse methods in image and signal processing, 13 hours, INSA Rennes, 5th year, Mathematical engineering, France.
- Master: C. Petit, Foundations of smart sensing, 13.5 hours, ENSAI, Master of Science in Statistics for Smart Data, France.
- Engineering degree: A. Roumy, Image and Video compression, 10 hours, University Rennes 1, ESIR, France.
- Master: A. Roumy, High dimensional statistical learning, 9 hours, University Rennes 1, SIF Master 2, France.

11.2.2 Supervision

- C. Guillemot is the PhD supervisor of Brandon Le Bon (IA chair DeepCIM, Oct. 2020-Sept. 2023), Rita Fermanian, IA chair DeepCIM, Oct. 2020-Sept. 2023, Davi Freitas (H2020 Marie Curie Plenoptima project, co-tutelle with Tampere University, Sept. 2021-Aug 2024), Samuel Willingham (H2020 Marie Curie Plenoptima project, co-tutelle with MidSweden University, June 2021-May 2024.
- C. Guillemot co-supervises two other PhD students recruited by Plenoptima partners, in a context of double degrees with Univ. of Rennes 1, at Tampere University (Ipek Anil Atalay, Aug. 2021-July 2024) and Midsweden University (Soheib Takhtardeshir, Jan. 2022-Dec. 2025).
- C. Guillemot and T. Maugey co-supervise Patrick Garus (Cifre contract with Orange lab., Jan.2019-Dec.2021), Kai Gu (H2020 Marie Curie Plenoptima project, co-tutelle with Technical University Berlin, June 2021-May 2024).
- C. Guillemot and A. Roumy co-supervise Pascal Bacchus and Arthur Lecert (Litchie project, Oct 2020-Sept. 2023), Yiqun Liu (Cifre contract with Ateme, Aug.2020-Jul.2023).

- T. Maugey is the PhD supervisor of Tom Bachard, PhD student, ministry grant.
- A. Roumy is the PhD supervisor of Nicolas Charpenay (ENS grant, Oct. 2020-Sept. 2023).
- A. Roumy and T. Maugey co-supervise Reda Kaafarani (Cifre contract with Mediakind, Dec.2020-April 2021), and Rémy Piau (CominLabs Colearn project, Sept. 2021- Aug. 2024).

11.2.3 Juries

- C. Guillemot was rapporteur of the PhD thesis of
  - Amélie Barbe, Univ. de Lyon / ENS- Lyon, 10 Dec. 2021.
- C. Guillemot was member of the PhD thesis of
  - Jinglei Shi, Univ. Rennes 1, 16 June, 2021.
  - Valentin Rebière, Sorbonne Univ., 30 June 2021
  - Simon Evain, Univ. Rennes 1, 12 July 2021
  - Guillaume Chataignier, Univ. Grenoble-Alpes, 18 Nov. 2021.
- T. Maugey was member of the PhD thesis of Muhammad Abeer Irfan, Politecnico di Torino, Italy (April 2021).
- A. Roumy has been reviewer of the PhD of:
  - Anthony Nasrallah, Univ. Paris Saclay, prepared at Telecom ParisTech, Dec. 2021
- A. Roumy chaired the PhD committee
  - Dadja Anade, Univ. Lyon, prepared at Insa Lyon, Oct. 2021
- A. Roumy has been member of the PhD committee of:
  - Pierre Stock, ENS Lyon, April 2021.
- A. Roumy served as a member of Board of Examiners (Comité de sélection)
  - for an Associate Professor position (Maitre de Conférences) at CREATIS Polytech Lyon (MCF61-776), June 2021.
  - for a Professor position (Professeur des Universités) at Engineering school ENSEA Cergy, June 2021.
- A. Roumy served as a jury member of the following recruitement competitons
  - Junior Researchers (CRCN/ISFP) at the Inria Sophia Antipolis center, May 2021
  - Senior Researchers (DR) at Inria, May 2021
  - Junior Researchers with disabilities (CRTH) at Inria, June 2021
- A. Roumy served as a member of the admission board (final step) of the Inria Senior Researchers' competition, June 2021.
- A. Roumy has been a member of the Inria Delegation committee of the Rennes Bretagne Atlantique center, Febr. 2021.

11.2.4 Internal or external Inria responsibilities

- T. Maugey is scientific mediation officer of the Inria Rennes Scientific mediation team.

11.2.5 Articles and contents

- C. Guillemot contributed to the joint paper [25] presenting the activities in artificial intelligence, at the IRISA and INRIA Bretagne atlantique centers, to the general public.

11.2.6 Interventions

- A. Roumy presented the research profession in order to promote research and create incentives, at INSA Rennes, “ parcours Recherche Innovation Entrepreneuriat", Dec. 2021.

12 Scientific production

12.1 Major publications

[1] R. Farrugia and C. Guillemot. ‘Light Field Super-Resolution using a Low-Rank Prior and Deep Convolutional Neural Networks’. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019), pp. 1–15. DOI: 10.1109/TPAMI.2019.2893666. URL: https://hal.archives-ouvertes.fr/hal-01984843.

[2] X. Jiang, M. Le Pendu, R. A. Farrugia and C. Guillemot. ‘Light Field Compression with Homography-based Low Rank Approximation’. In: *IEEE Journal of Selected Topics in Signal Processing* (2017). DOI: 10.1109/JSTSP.2017.2747078. URL: https://hal.archives-ouvertes.fr/hal-01591349.

[3] M. Le Pendu, C. Guillemot and A. Smolic. ‘A Fourier Disparity Layer representation for Light Fields’. In: *IEEE Transactions on Image Processing* (May 2019), pp. 5740–5753. DOI: 10.1109/TIP.2019.2922099. URL: https://hal.archives-ouvertes.fr/hal-02130555.

[4] N. Mahmoudian Bidgoli, T. Maugey and A. Roumy. ‘Fine granularity access in interactive compression of 360-degree images based on rate-adaptive channel codes’. In: *IEEE Transactions on Multimedia* (2020). DOI: 10.1109/TMM.2020.3017890. URL: https://hal.inria.fr/hal-02946795.

[5] M. Q. Pham, A. Roumy, T. Maugey, E. Dupraz and M. Kieffer. ‘Optimal Reference Selection for Random Access in Predictive Coding Schemes’. In: *IEEE Transactions on Communications* 68.9 (2020), pp. 5819–5833. DOI: 10.1109/TCOMM.2020.3002937. URL: https://hal-imt-atlantique.archives-ouvertes.fr/hal-02925113.

[6] M. Rizkallah, X. Su, T. Maugey and C. Guillemot. ‘Geometry-Aware Graph Transforms for Light Field Compact Representation’. In: *IEEE Transactions on Image Processing* (Aug. 2019), pp. 1–15. DOI: 10.1109/TIP.2019.2928873. URL: https://hal.archives-ouvertes.fr/hal-02199839.

12.2 Publications of the year

International journals

[7] P. Garus, F. HENRY, J. Jung, T. Maugey and C. Guillemot. ‘Immersive Video Coding: Should Geometry Information be Transmitted as Depth Maps?’ In: *IEEE Transactions on Circuits and Systems for Video Technology* (July 2021), pp. 1–15. URL: https://hal.archives-ouvertes.fr/hal-03303040.

[8] F. Hawary, G. Boisson, C. Guillemot and P. Guillotel. ‘Compressively Sampled Light Field Reconstruction Using Orthogonal Frequency Selection and Refinement’. In: *Signal Processing: Image Communication* 92 (Mar. 2021), p. 116087. DOI: 10.1016/j.image.2020.116087. URL: https://hal.archives-ouvertes.fr/hal-03028645.

[9] M. Krivokuća, E. Miandji, C. Guillemot and P. A. Chou. ‘Compression of Plenoptic Point Cloud Attributes Using 6-D Point Clouds and 6-D Transforms’. In: *IEEE Transactions on Multimedia* (2021), pp. 1–15. DOI: 10.1109/TMM.2021.3129341. URL: https://hal.archives-ouvertes.fr/hal-03432597.

[10] G. Le Guludec, E. Miandji and C. Guillemot. 'Deep Light Field Acquisition Using Learned Coded Mask Distributions for Color Filter Array Sensors'. In: *IEEE Transactions on Computational Imaging* 7 (May 2021), pp. 475–488. DOI: 10.1109/TCI.2021.3077131. URL: https://hal.archives-ouvertes.fr/hal-03203347.

[11] E. Miandji, H.-N. Nguyen, S. Hajisharif, J. Unger and C. Guillemot. 'Compressive HDR light field imaging using a single multi-ISO sensor'. In: *IEEE Transactions on Computational Imaging* (Dec. 2021), pp. 1–16. URL: https://hal.archives-ouvertes.fr/hal-03456792.

[12] H. N. Nguyen, E. Miandji and C. Guillemot. 'Multi-Mask Camera Model for Compressed Acquisition of Light Fields'. In: *IEEE Transactions on Computational Imaging* 7 (2021), pp. 191–208. DOI: 10.1109/TCI.2021.3053702. URL: https://hal.archives-ouvertes.fr/hal-03104409.

[13] M. Rizkallah, T. Maugey and C. Guillemot. 'Rate-Distortion Optimized Graph Coarsening and Partitioning for Light Field Coding'. In: *IEEE Transactions on Image Processing* (May 2021), pp. 1–14. URL: https://hal.archives-ouvertes.fr/hal-03230325.

[14] Z. Xiao, J. Shi, X. Jiang and C. Guillemot. 'A learning-based view extrapolation method for axial super-resolution'. In: *Neurocomputing* (May 2021), pp. 1–13. URL: https://hal.archives-ouvertes.fr/hal-03230321.

[15] F. Ye, N. M. Bidgoli, E. Dupraz, A. Roumy, K. Amis and T. Maugey. 'Bit-Plane Coding in Extractable Source Coding: Optimality, Modeling, and Application to 360° Data'. In: *IEEE Communications Letters* 25.5 (May 2021), pp. 1412–1416. DOI: 10.1109/LCOMM.2021.3050932. URL: https://hal-imt-atlantique.archives-ouvertes.fr/hal-03294082.

International peer-reviewed conferences

[16] N. Charpenay, M. Le Treust and A. Roumy. 'Zero-error source coding when side information may be present'. In: International Zurich Seminar on Information and Communication. Zurich, Switzerland, 2021. URL: https://hal.archives-ouvertes.fr/hal-03290860.

[17] S. Evain and C. Guillemot. 'A Neural Network with Adversarial Loss for Light Field Synthesis from a Single Image'. In: VISAPP 2021 - 16th International Conference on Computer Vision Theory and Applications. Vienna, Austria, Feb. 2021, pp. 1–10. URL: https://hal.archives-ouvertes.fr/hal-03024210.

[18] R. Fermanian, M. Le Pendu and C. Guillemot. 'Regularizing the Deep Image Prior with a Learned Denoiser for Linear Inverse Problems'. In: MMSP 2021 - IEEE 23rd International Workshop on Multimedia Siganl Processing. Tampere, Finland: IEEE, 6th Oct. 2021, pp. 1–6. URL: https://hal.archives-ouvertes.fr/hal-03310533.

[19] R. Kaafarani, M. Blestel, T. Maugey, M. Ropert and A. Roumy. 'Evaluation Of Bitrate Ladders For Versatile Video Coder'. In: VCIP 2021 - IEEE Visual Communications and Image Processing. Munich, Germany, 5th Dec. 2021, pp. 1–5. URL: https://hal.inria.fr/hal-03483326.

[20] A. Marie, N. Mahmoudian Bidgoli, T. Maugey and A. Roumy. 'Rate-distortion optimized motion estimation for on-the-sphere compression of 360 videos'. In: ICASSP 2021 - IEEE International Conference on Acoustics, Speech and Signal Processing. Toronto, Canada, 6th June 2021, pp. 1–5. URL: https://hal.inria.fr/hal-03484164.

[21] J. Shi, X. Jiang and C. Guillemot. 'Deep Video Frame Rate Up-conversion Network using Feature-based Progressive Residue Refinement a'. In: International Conference on Computer Vision Theory and Applications. online, France, Feb. 2022. URL: https://hal.archives-ouvertes.fr/hal-03432380.

[22] Z. Xiao, M. Eng Zhang, H. Jin and C. Guillemot. 'A light field FDL-HSIFT feature in scale-disparity space'. In: ICIP 2021 - IEEE International Conference on Image Processing. Anchorage, United States, 19th Sept. 2021, pp. 1–5. URL: https://hal.archives-ouvertes.fr/hal-03233522.

Doctoral dissertations and habilitation theses

[23] S. Evain. 'Deep learning for light field view synthesis from monocular and a very sparse set of input views'. Université Rennes 1, 12th July 2021. URL: https://hal.archives-ouvertes.fr/tel-03428769.

[24] J. Shi. 'Learning-based depth estimation from light field and view synthesis'. Université Rennes 1, 16th June 2021. URL: https://hal.archives-ouvertes.fr/tel-03270446.

Other scientific publications

[25] G. Gravier, E. Fromont, N. Courty, T. Furon, C. Guillemot and P. Robuffo Giordano. 'Rennes - une IA souveraine au service de la vie publique'. In: *Bulletin de l'Association Française pour l'Intelligence Artificielle* (2021). URL: https://hal.archives-ouvertes.fr/hal-03313161.